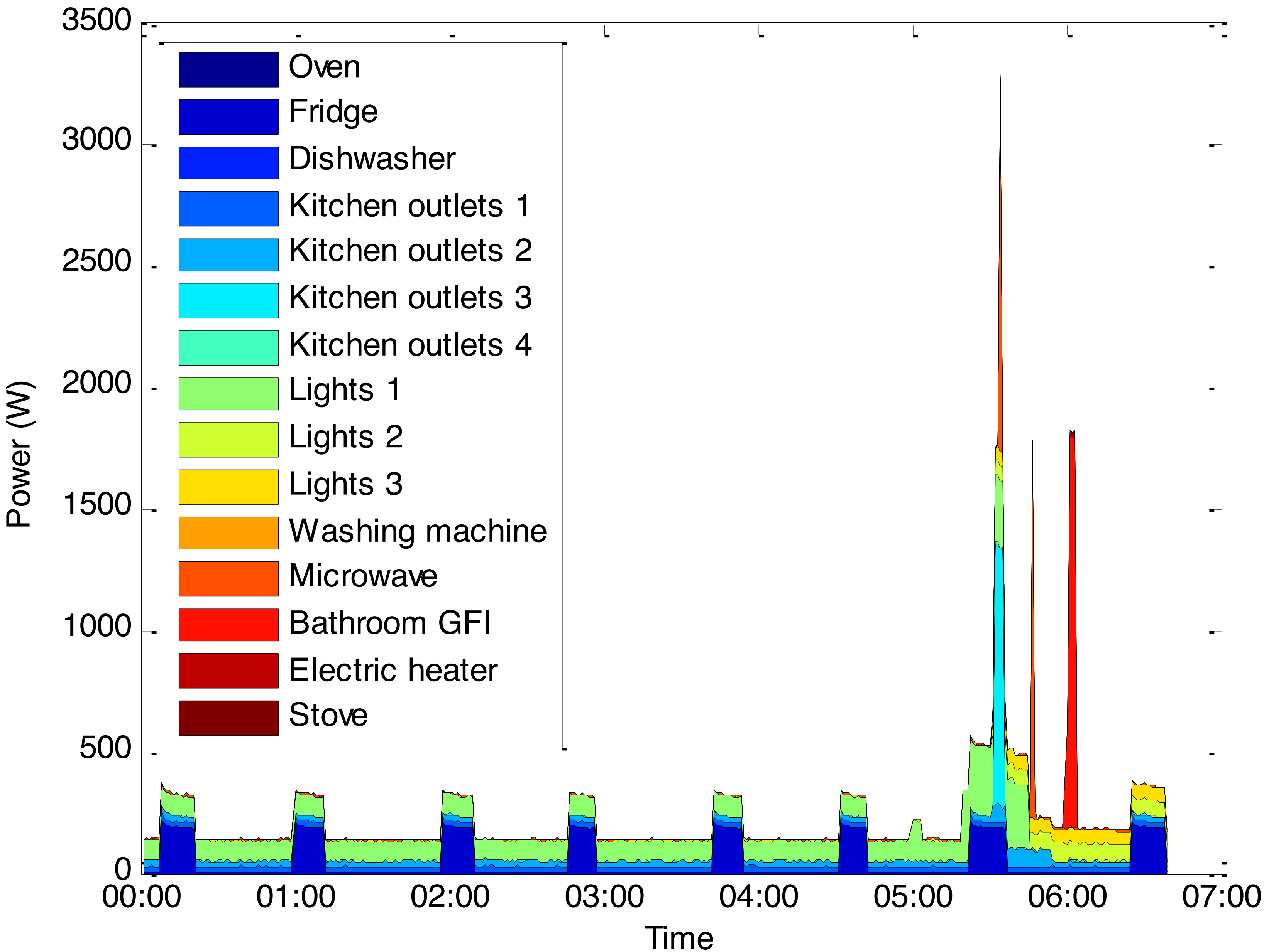
Dataport and NILMTK: A Building Data Set Designed for Non-intrusive Load Monitoring

Oliver Parson, Grant Fisher, April Hersey, Nipun Batra, Jack Kelly, Amarjeet Singh, William Knottenbelt, Alex Rogers

osp@ecs.soton.ac.uk

blog.oliverparson.co.uk


Oven
Fridge
Dishwasher
Kitchen outlets 1
Kitchen outlets 2
Kitchen outlets 3
Kitchen outlets 4
Lights 1
Lights 2
Lights 3
Washing machine
Microwave
Bathroom GFI
Electric heater
Stove
Power (W)
3500
3000
2500
2000
1500
1000
500
0
00:00
01:00
02:00
03:00
04:00
05:00
06:00
07:00
Time

Approximate Inference in Additive Factorial HMMs with Application to Energy Disaggregation

Tommi Jaakkola
MIT CSAIL
tommi@csail.mit.edu

Transient Event Detection for Nonintrusive Load Monitoring and Demand Side Management Using Voltage Distortion

Robert Cox and Steven B. Leeb
Laboratory for Electromagnetic and Electronic Systems
Massachusetts Institute of Technology
Cambridge, MA 02139
Email: rwcox@mit.edu, sbleeb@mit.edu

Steven R. Shaw
Department of Electrical and Computer Engineering
Montana State University
Bozeman, MT

Leslie K. Norford
Department of Architecture
Massachusetts Institute of Technology
Cambridge, MA 02139
Email: lnorford@mit.edu

Abstract—**This paper describes a simple system that can be used for autonomous demand-side management in a load site such as a home or commercial facility. The system identifies the operation of individual loads using transient patterns observed in the voltage waveform measured at an electric service outlet. The theoretical foundation of the measurement process is introduced, and a preprocessor that computes short-time estimates of the spectral content of the voltage waveform is described. The paper presents several example measurements demonstrating the ability of the system to obtain estimates of the spectral content of the voltage waveform.**

I. Introduction

Recent estimates indicate that products containing power supplies consume over 200 billion kWh of energy each year in the United States [1]. Additionally, market projections suggest that this number will grow over time, as worldwide sales of power supplies are expected to grow approximately 15% each year [1]. With so much energy consumed by power electronic circuits, it is clear that one way to effect a significant reduction in peak energy consumption at a load site such as a home or commercial facility is to introduce autonomous demand-side energy management features. For example, the control circuitry included in switching power supplies could incorporate a consumption control feature that would place the device in a low power mode following the detection of the operation of a large energy consumer such as an electric water heater. This paper describes a potentially inexpensive system that could allow a device to sense the operation of other loads on the local utility distribution network using measurements of the local utility voltage.

The method used here to identify the operation of individual loads is based on the observation that the transient behavior of an electrical load is strongly influenced by the task that the load performs [2]. Consequently, different classes of loads posses unique and repeatedly observable transient profiles that can serve as "fingerprints" indicating the operation of individual loads. For instance, the turn-on transients associated with an incandescent lamp and an induction motor are distinctly different, as the physical task of heating the cold filament of a lamp is not the same as accelerating a rotor [2]. That concept has been used to develop a device known as the nonintrusive load monitor (NILM) that can detect the operation of individual loads using transient patterns observed in the short-time estimates of the spectral content of the aggregate current drawn by a collection of loads [2], [3]. The prototype system described in this paper takes the nonintrusive monitoring concept one step further by identifying load operation using the voltage distortion caused by transient load currents.

The paper begins in Section II by introducing the concepts that motivate nonintrusive load monitoring and by describing the theoretical background upon which the current system is based. Section III describes both the continuous-time and discrete-time operations performed by the preprocessor that computes estimates of the spectral content of the measured voltage waveform. Additionally, Section III also provides numerous examples that demonstrate the ability of the preprocessor to sense the small changes in the voltage waveform that are the result of individual load currents. Section IV provides a description of the methods used by the prototype system to identify the operation of individual loads, and it also includes several examples that demonstrate the success of the load identification scheme. Finally, the paper concludes in Section V by summarizing the results and by describing several areas of ongoing research.

II. Spectral Envelope Estimation

In standard approaches to nonintrusive monitoring, loads are detected using the two step procedure shown in Fig. 1 [2], [3]. In the first step, a preprocessor, which can be implemented using either analog electronics or digital software routines, computes estimates of the short-time spectral content of a measured current waveform [2]–[4]. Subsequently, the spectral estimates created by the preprocessor are passed to a software module that identifies the operation of individual loads by matching stored templates to transient patterns observed in the preprocessed data stream [3].

0-7803-9547-6/06/$20.00 ©2006 IEEE.

...sufficient for massive implementation of NIALM [4].

For the data sampling rates attainable with smart grids (up to 1 Hz), the established NIALM algorithms are based on matching the observed step-wise power changes with appliances being turned on or off [3], [5]. This matching is prone to both measurement and algorithmic errors and to the ambiguity related to a simultaneous start or end of multiple appliances. The main reason, nonetheless, for the monitoring accuracy being low is an overlap in the power draw between different appliances [4].

In this paper, we propose a new NIALM algorithm capable of dramatic improvement of disaggregation accuracy. In the proposed algorithm, the matching between the measurable power observations and the appliances is made by consideration of a series of transitions among the appliance

978-1-4244-9894-9/11/$26.00 ©2011 IEEE

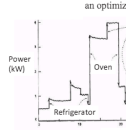


Fig. 1 Power vs. time and corresponding ...

Fig. 1 illustrates step 1 of this method. For the appliances considered, the 2D clusters are quite distinct; however, there is no guarantee that the clusters do not overlap in a general case. Also, it is unlikely that power-grid meters will provide both

Bayesian Nonparametric Hidden Semi-Markov Models

Matthew J. Johnson MATTJJ@CSAIL.MIT.EDU
Alan S. Willsky WILLSKY@MIT.EDU
Laboratory for Information and Decision Systems
Department of EECS
Massachusetts Institute of Technology
Cambridge, MA 02139-4307, USA

Abstract

There is much interest in the Hierarchical Dirichlet Process Hidden Markov Model (HDP-HMM) as a natural Bayesian nonparametric extension of the ubiquitous Hidden Markov Model for learning from sequential and time-series data. However, in many settings the HDP-HMM's strict Markovian constraints are undesirable, particularly if we wish to learn or encode non-geometric state durations. We can extend the HDP-HMM to capture such structure by drawing upon explicit-duration semi-Markov modeling, which has been developed mainly in the parametric non-Bayesian setting, to allow construction of highly interpretable models that admit natural prior information on state durations.

In this paper we introduce the explicit-duration Hierarchical Dirichlet Process Hidden semi-Markov Model (HDP-HSMM) and develop sampling algorithms for efficient posterior inference. The methods we introduce also provide new methods for sampling inference in the finite Bayesian HSMM. Our modular Gibbs sampling methods can be embedded in samplers for larger hierarchical Bayesian models, adding semi-Markov chain modeling as another tool in the Bayesian inference toolbox. We demonstrate the utility of the HDP-HSMM and our inference methods on both synthetic and real experiments.

Keywords: Bayesian nonparametrics, time series, semi-Markov, sampling algorithms, Hierarchical Dirichlet Process Hidden Markov Model

1. Introduction

Given a set of sequential data in an unsupervised setting, we often aim to infer meaningful states, or "topics," present in the data along with characteristics that describe and distinguish those states. For example, in a speaker diarization (or who-spoke-when) problem, we are given a single audio recording of a meeting and wish to infer the number of speakers present, when they speak, and some characteristics governing their speech patterns (Tranter and Reynolds, 2006; Fox et al., 2008). Or in separating a home power signal into the power signals of individual devices, we would be able to perform the task much better if we were able to exploit our prior knowledge about the levels and durations of each device's power modes (Kolter and Johnson, 2011). Such learning problems for sequential data are pervasive, and so we would like to build general models that are both flexible enough to be applicable to many domains and expressive enough to encode the appropriate information.

Hidden Markov Models (HMMs) have proven to be excellent general models for approaching learning problems in sequential data, but they have two significant disadvantages: (1) state duration distributions are necessarily restricted to a geometric form that is not

...the United States, electricity constitutes 38% of all energy used, and residential and commercial buildings together use 75% of this electricity [28]; thus, this 12% figure accounts for a sizable amount of energy that could potentially be saved. However, the widely-available sensors that provide electricity consumption information, namely the so-called "Smart Meters" that are already becoming ubiquitous, collect energy information only at the whole-home level and at a very low resolution (typically every hour or 15 minutes). Thus, energy disaggregation methods that can take this whole-home data and use it to predict individual appliance usage present an algorithmic challenge where advances can have a significant impact on large-scale energy efficiency issues.

1

Extracting Features from an Electrical Signal of a Non-Intrusive Load Monitoring System

Marisa B. Figueiredo[1], Ana de Almeida[1], Bernardete Ribeiro[1], and António Martins[2]

[1] CISUC - Center for Informatics and Systems, University of Coimbra, Polo II, P-3030-290 Coimbra, Portugal
mbfig@dei.uc.pt, amca@mat.uc.pt, bribeiro@dei.uc.pt
[2] ISA - Intelligent Sensing Anywhere, S.A., Rua D. Manuel I, P- 3030-320 Coimbra, Portugal
amartins@isa.pt

Abstract. Improving energy efficiency by monitoring household electrical consumption is of significant importance with the present-day climate change concerns. A solution for the electrical consumption management problem is the use of a non-intrusive load monitoring system (NILM). This system captures the signals from the aggregate consumption, extracts the features from these signals and classifies the extracted features in order to identify the switched on appliances. An effective device identification (ID) requires a signature to be assigned for each appliance. Moreover, to specify an ID for each device, signal processing techniques are needed for extracting the relevant features. This paper describes a technique for the steady-states recognition in an electrical digital signal as the first stage for the implementation of an innovative NILM. Furthermore, the final goal is to develop an intelligent system for the identification of the appliances by automated learning. The proposed approach is based on the ratio value between rectangular areas defined by the signal samples. The computational experiments show the method effectiveness for the accurate steady-states identification in the electrical input signals.

Keywords: Automated learning and identification, feature extraction and classification, non-intrusive load monitoring.

Introduction

...veral concepts that have recently arisen with the idea of Smart Environments ... for an application able to accurately identify and monitor electrical appli... ...es consumptions, like Smart Grids or in-Home Activity Tracking. Further... ...re, the monitoring systems must be inconspicuous. The use of ubiquitous ...nputing to develop smart systems by designing a non-intrusive load mon... ...ring system (NILM) satisfies all these requirements. Although the idea of ...NILM system dates from the eighties[1], only today could it achieve its full

...The Electric Power Research Institute sponsored the research on NILM systems, ...hich resulted in the American patent number 4858141, approved in 1989.

...Fyfe et al. (Eds.): IDEAL 2010, LNCS 6283, pp. 210–217, 2010.
...Springer-Verlag Berlin Heidelberg 2010

nilmtk / **nilmtk**

Unwatch 28 | Unstar 100 | Fork 65

Code | Issues 123 | Pull requests 0 | Wiki | Pulse | Graphs | Settings

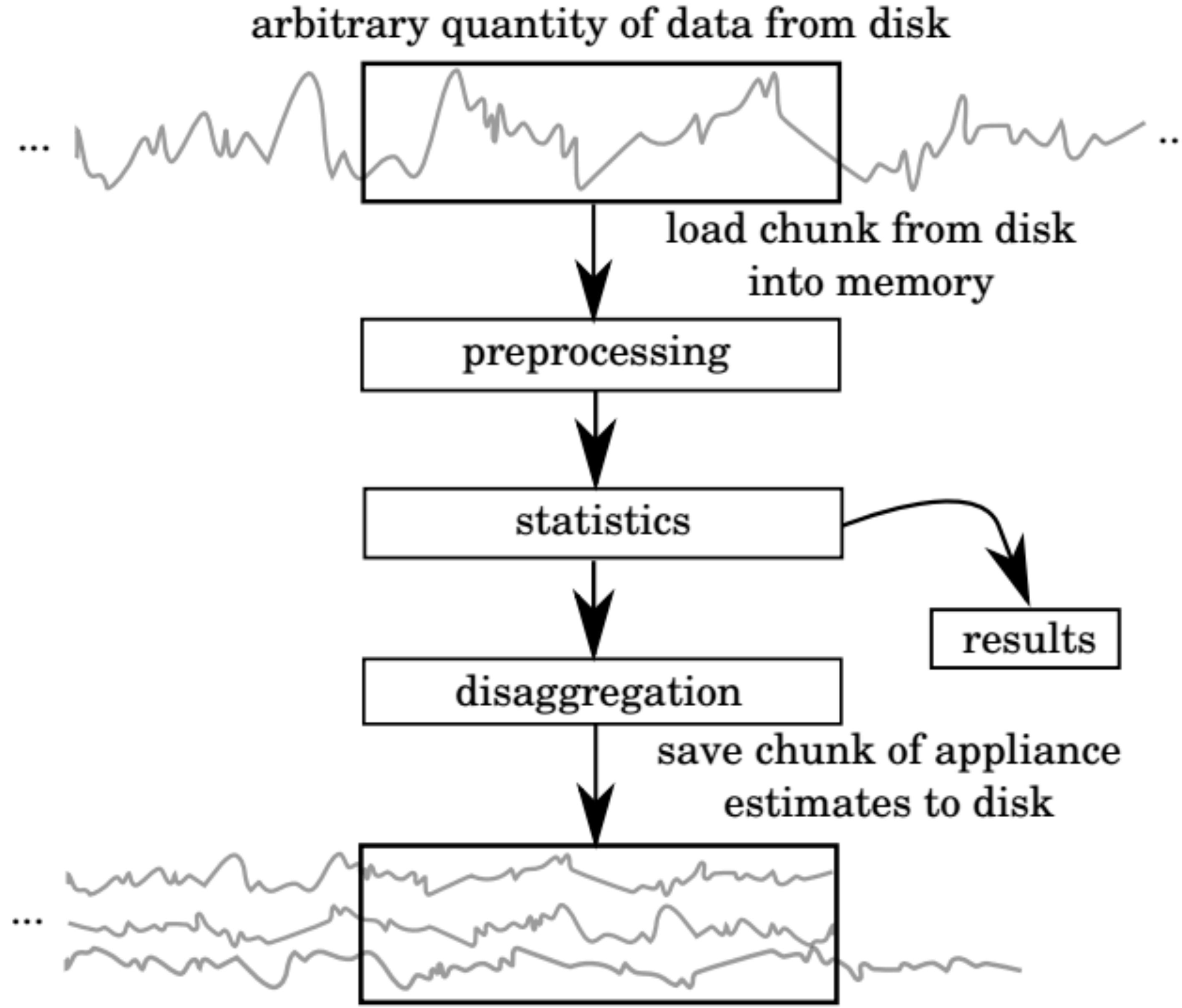
Non-Intrusive Load Monitoring Toolkit (nilmtk) http://nilmtk.github.io — Edit

1,588 commits | **1** branch | **4** releases | **12** contributors

Branch: **master** | **New pull request** | New file | Find file | HTTPS | https://github.com/nilmtk/r | Download ZIP

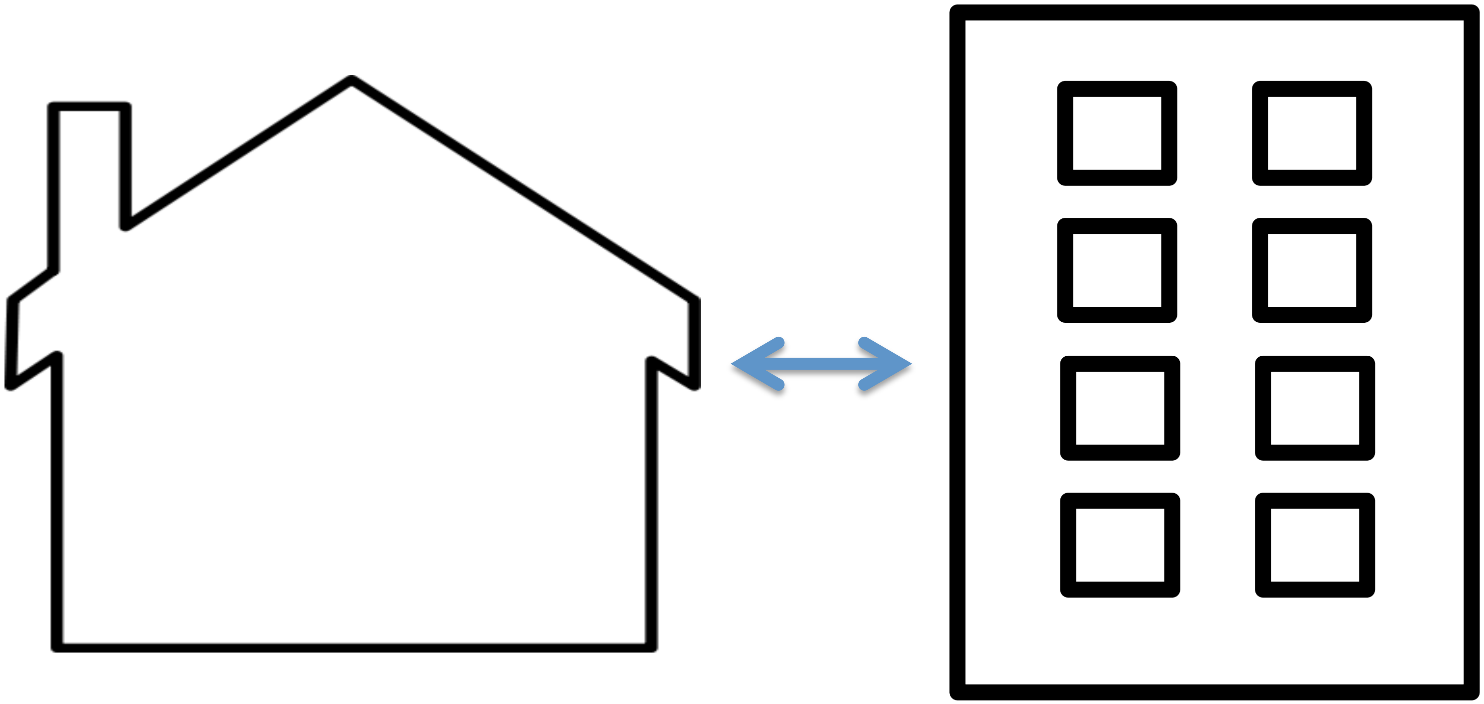
JackKelly remove `sh` dependency. Not needed #462 — Latest commit 064b4ed 5 days ago

data	untracking files from REDD and ukdale datasets	6 months ago
docs	fix links again	17 days ago
nilmtk	Added DRED datalink	5 days ago
notebooks	Added file for #449	a month ago
tests_on_large_datasets	Now iAWE matches v 0.1	6 months ago
.coveragerc	Added coverage testing..	a year ago
.gitignore	after merging from new_disag branch	6 months ago
.travis.yml	duh! do conda install after setting path!	17 days ago
LICENSE	Initial commit	2 years ago
README.md	capitalised NILMTK	2 months ago


arbitrary quantity of data from disk
...
...
load chunk from disk
into memory
preprocessing
statistics
results
disaggregation
save chunk of appliance
estimates to disk
...

Data set	**Number of houses**
REDD (2011)	6
BLUED (2012)	1
Smart* (2012)	3
HES (2012)	251
AMPds 2 (2013)	1
iAWE (2013)	1
UK-DALE (2014)	5
ECO (2014)	6
GREEND (2014)	9
SustData (2014)	50
DRED (2015)	1

Not all houses were created equally…



Powered by Pecan Street's energy data analytics and its multi-state research network, Sol is a free service that diagnoses potential maintenance issues with your home's solar PV system and provides daily updates on the amount and value of your system's generation.

PECAN STREET

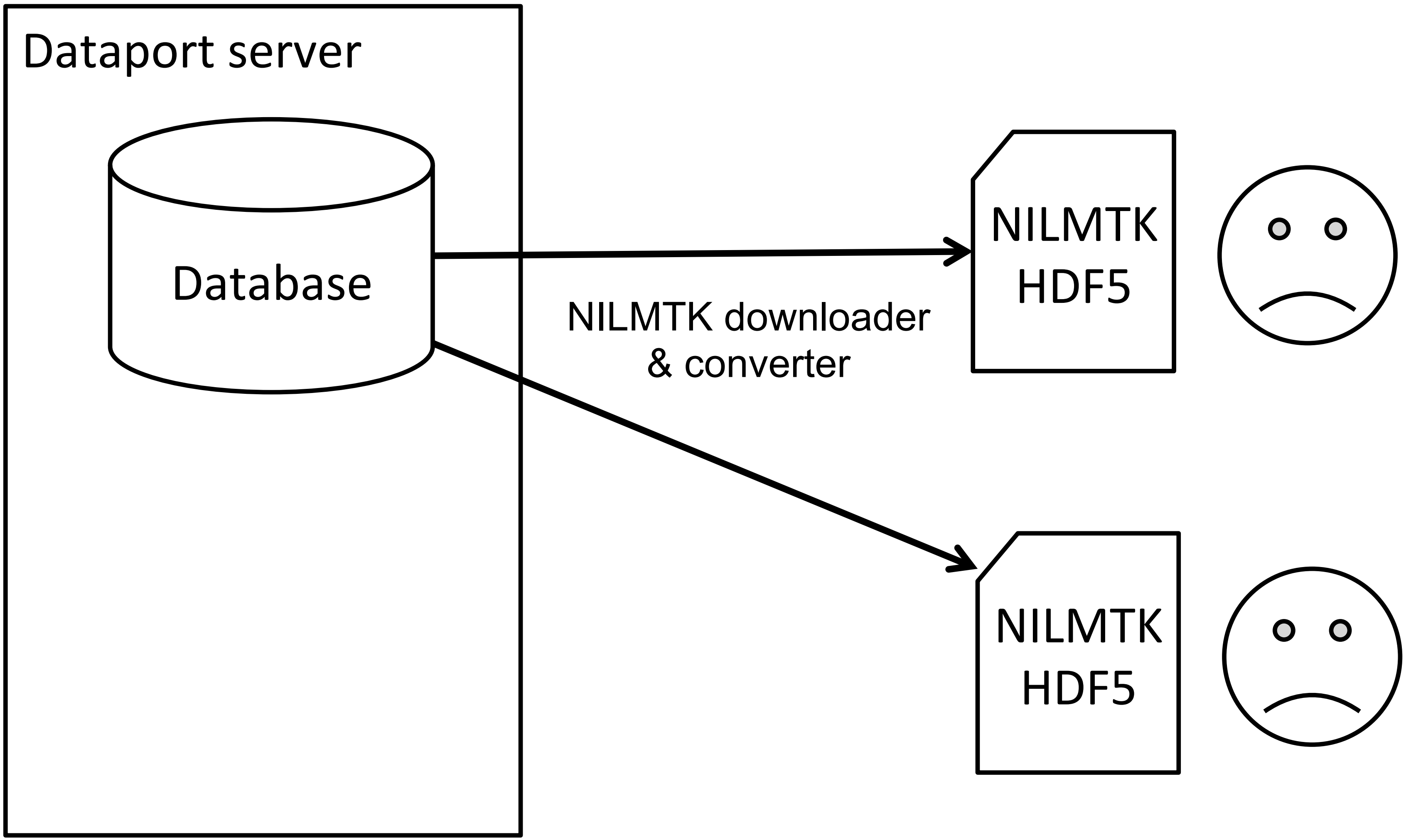
The World's Largest Energy Data Resource

Dataport is the world's largest source of disaggregated customer energy data for university researchers around the world. Access is free for university researchers, but registration and approval are required. Please log in or sign up.

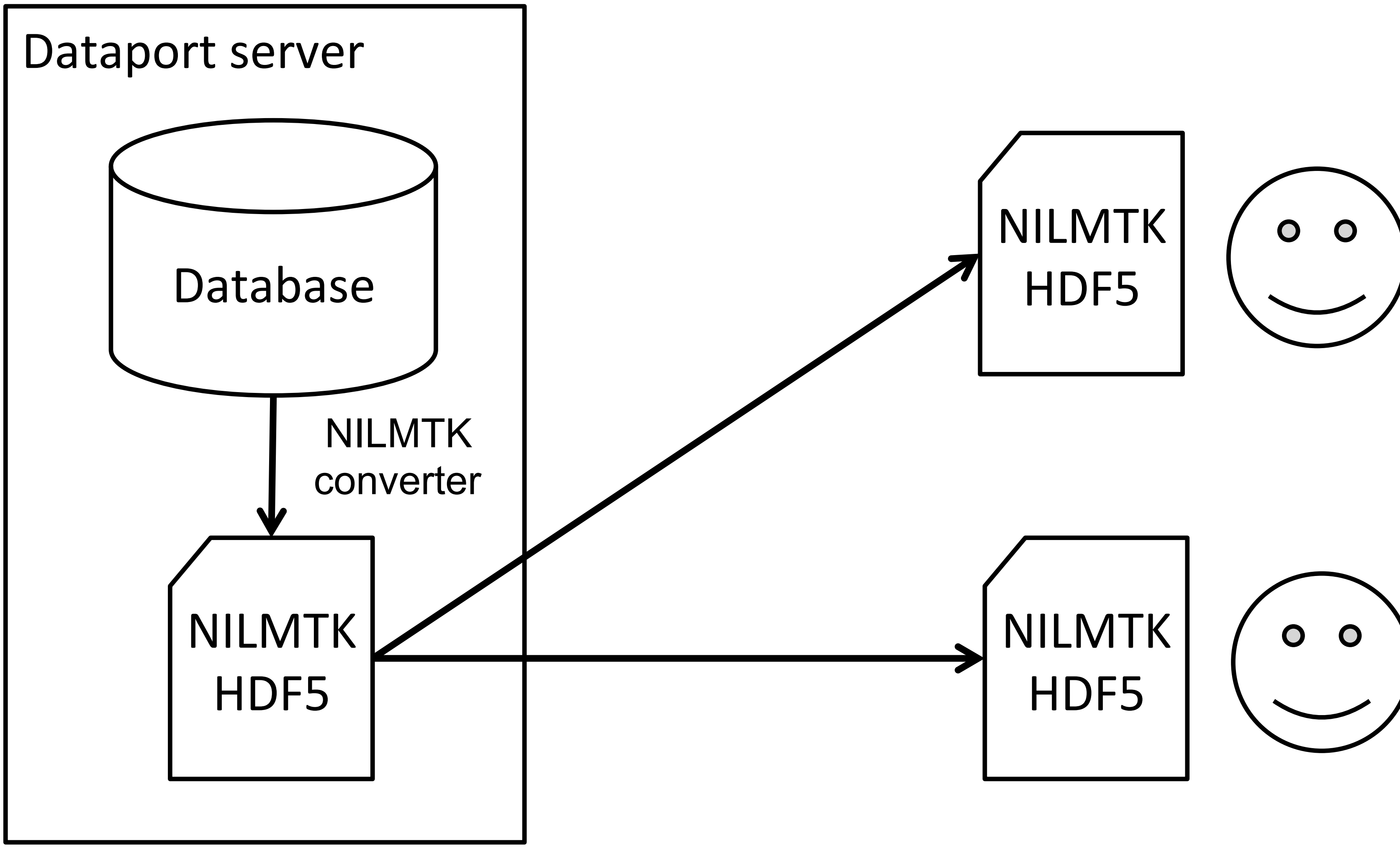
Dataport is divided into two main sections, both of which are accessible from the main navigation. The Knowledge Base houses reports and data visualizations developed by Pecan Street and other researchers, as well as industry job postings and the Pecan Street's blog posts. The Data section includes details about Pecan Street's data and allows researchers to query Pecan Street's energy database, create custom visualizations and download datasets. You can also bookmark this page -- new visualizations and reports will be posted here first.

If you have any questions, please email us at info@pecanstreet.org.

Old way


Dataport server
Database
NILMTK downloader
& converter
NILMTK
HDF5
NILMTK
HDF5

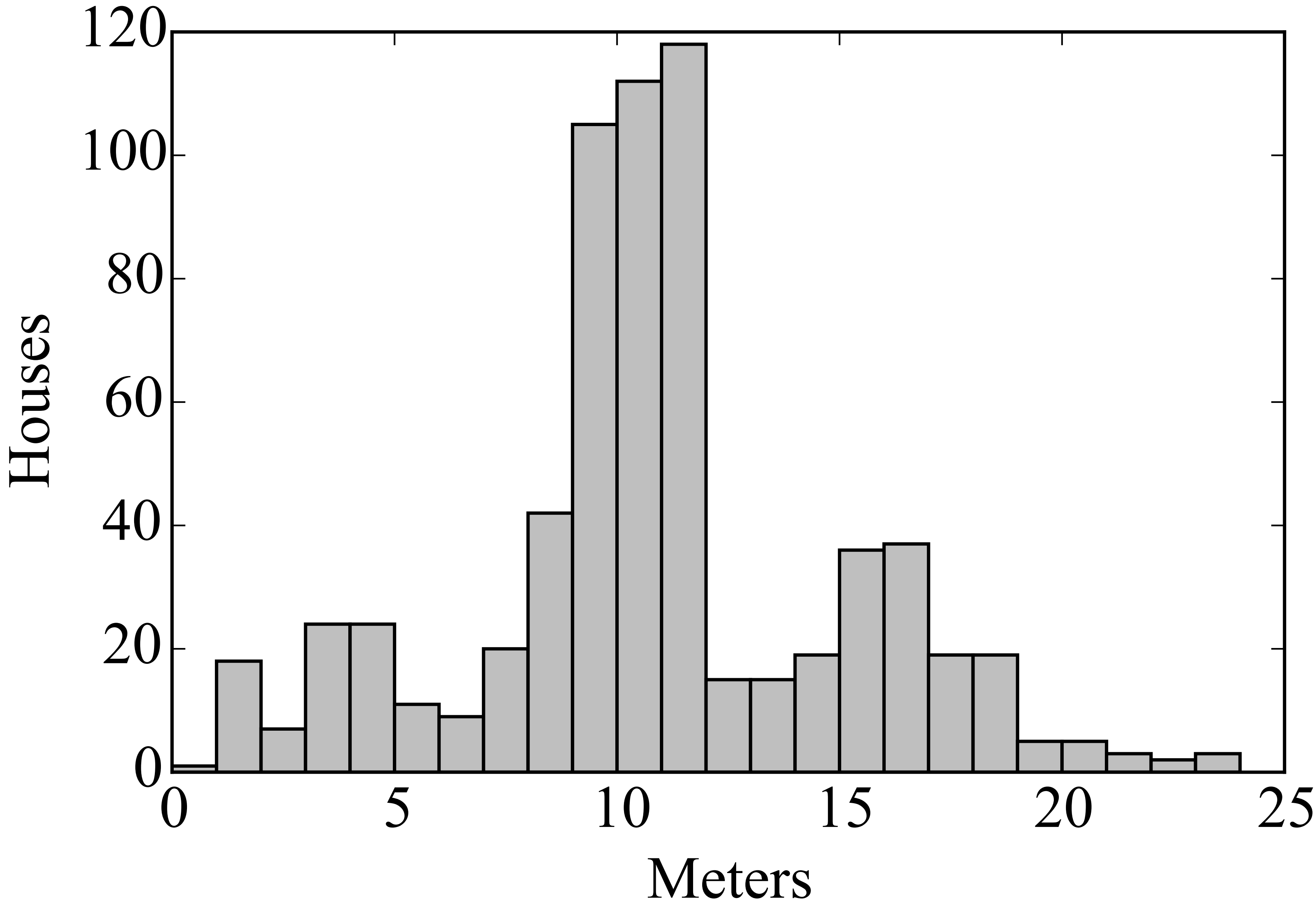
New way



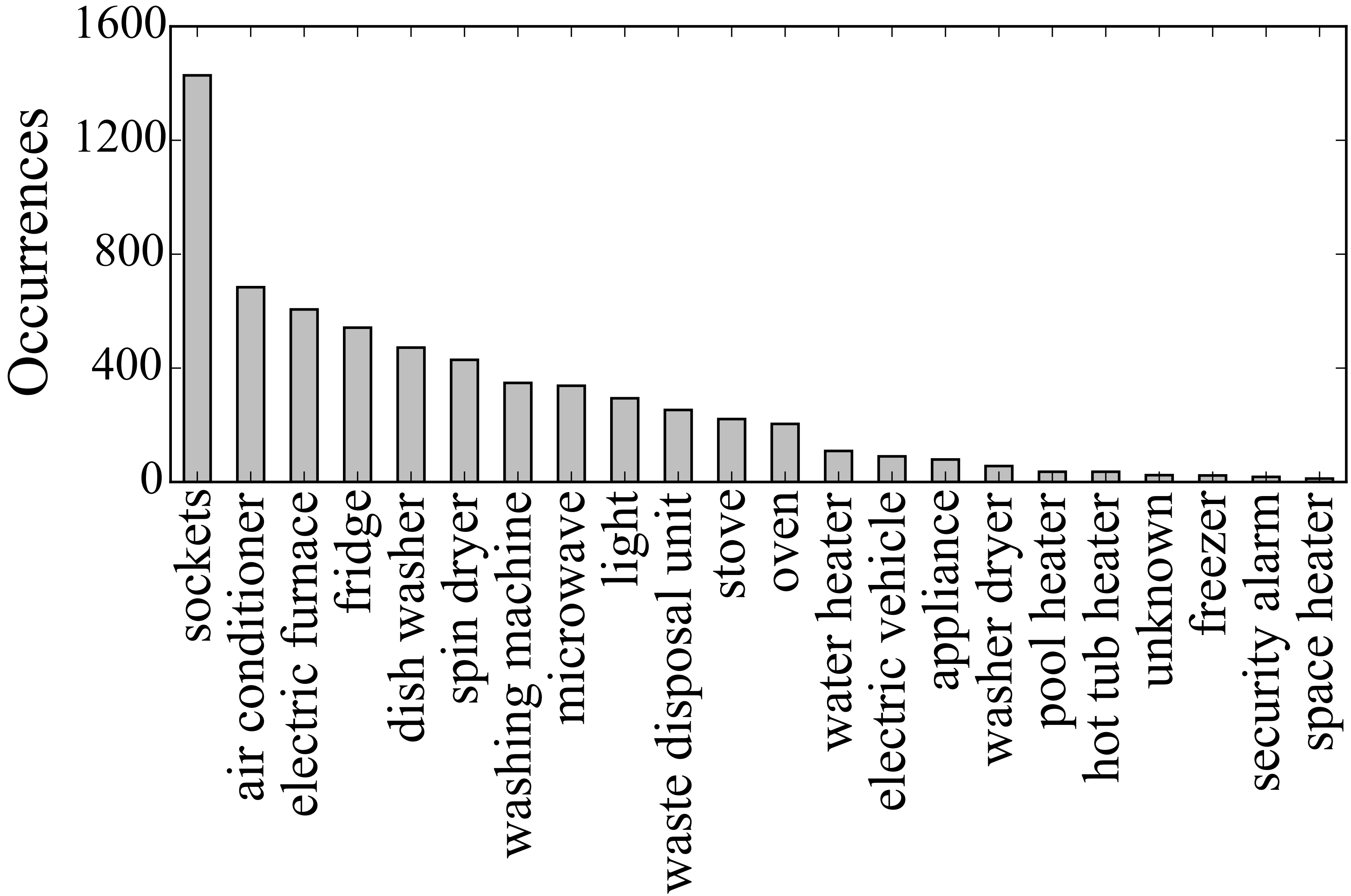
Dataport server
Database
NILMTK
converter
NILMTK
HDF5
NILMTK
HDF5
NILMTK
HDF5

The Dataport HDF5 file

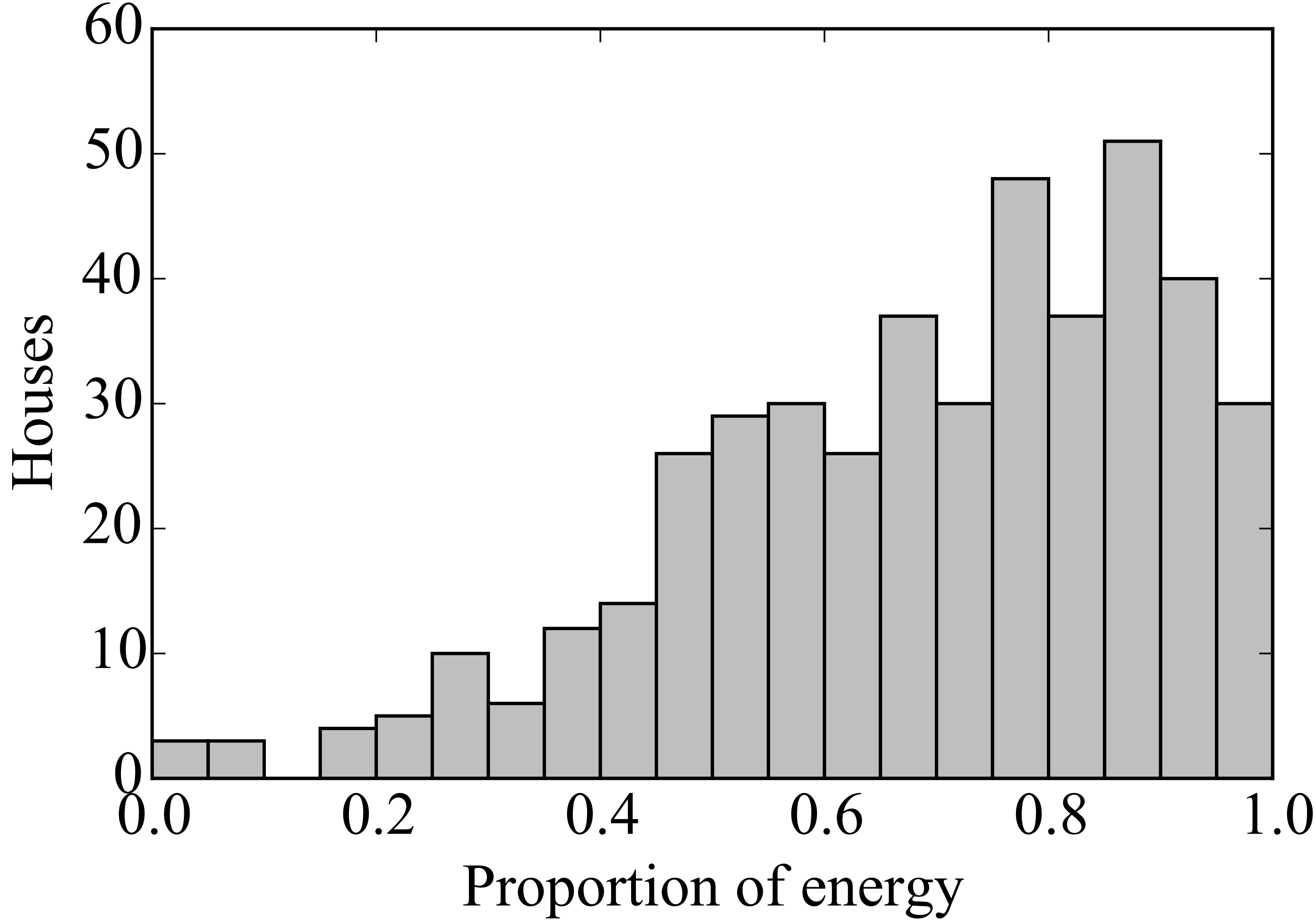
- 669 households
- Up to 23 circuit-level meters per household
- 1 measurement per minute
- 1 month data per household
- 1 GB in size



Houses
Meters
0
20
40
60
80
100
120
0
5
10
15
20
25

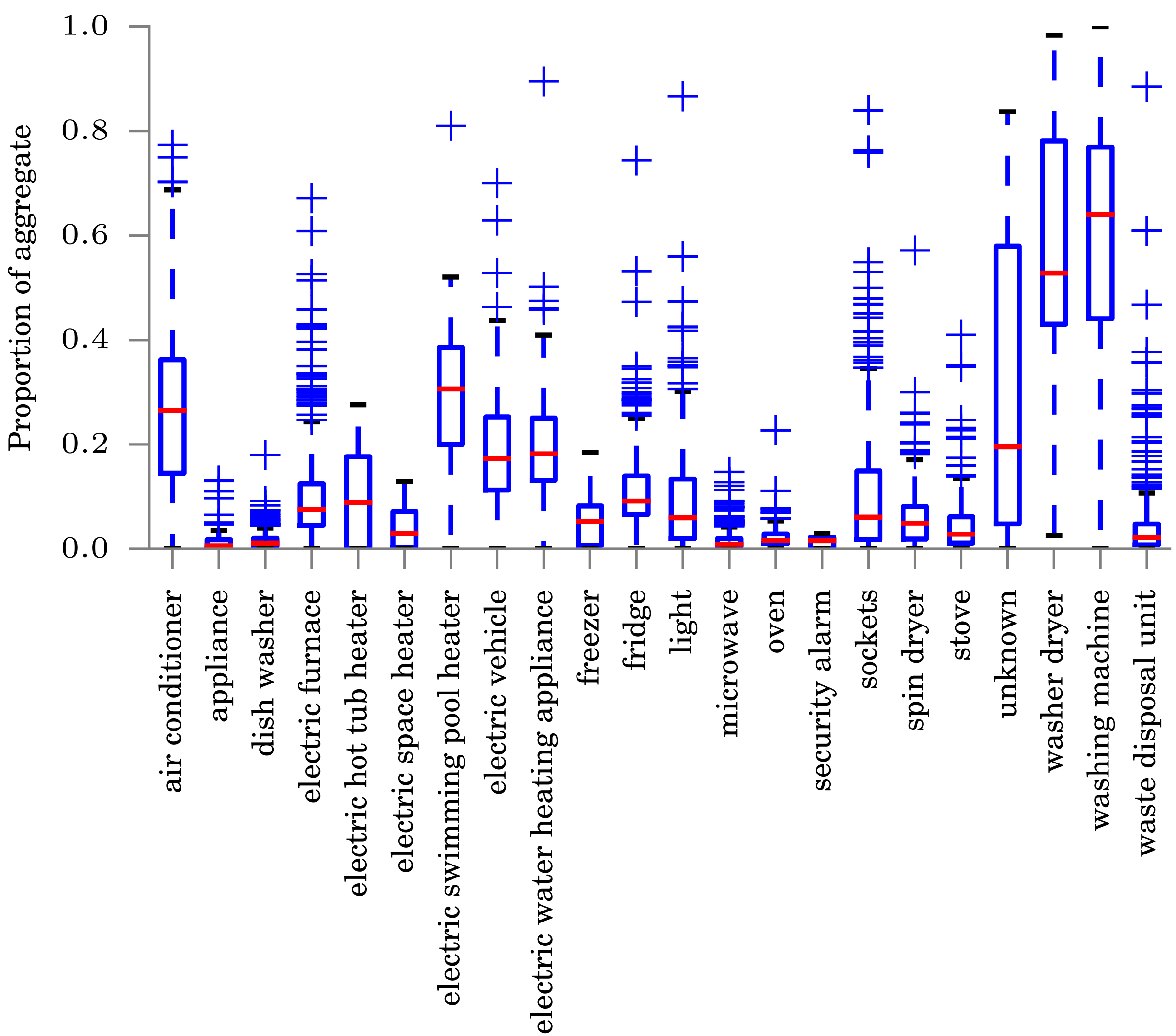


1600
1200
800
400
0
Occurrences
sockets
air conditioner
electric furnace
fridge
dish washer
spin dryer
washing machine
microwave
light
waste disposal unit
stove
oven
water heater
electric vehicle
appliance
washer dryer
pool heater
hot tub heater
unknown
freezer
security alarm
space heater

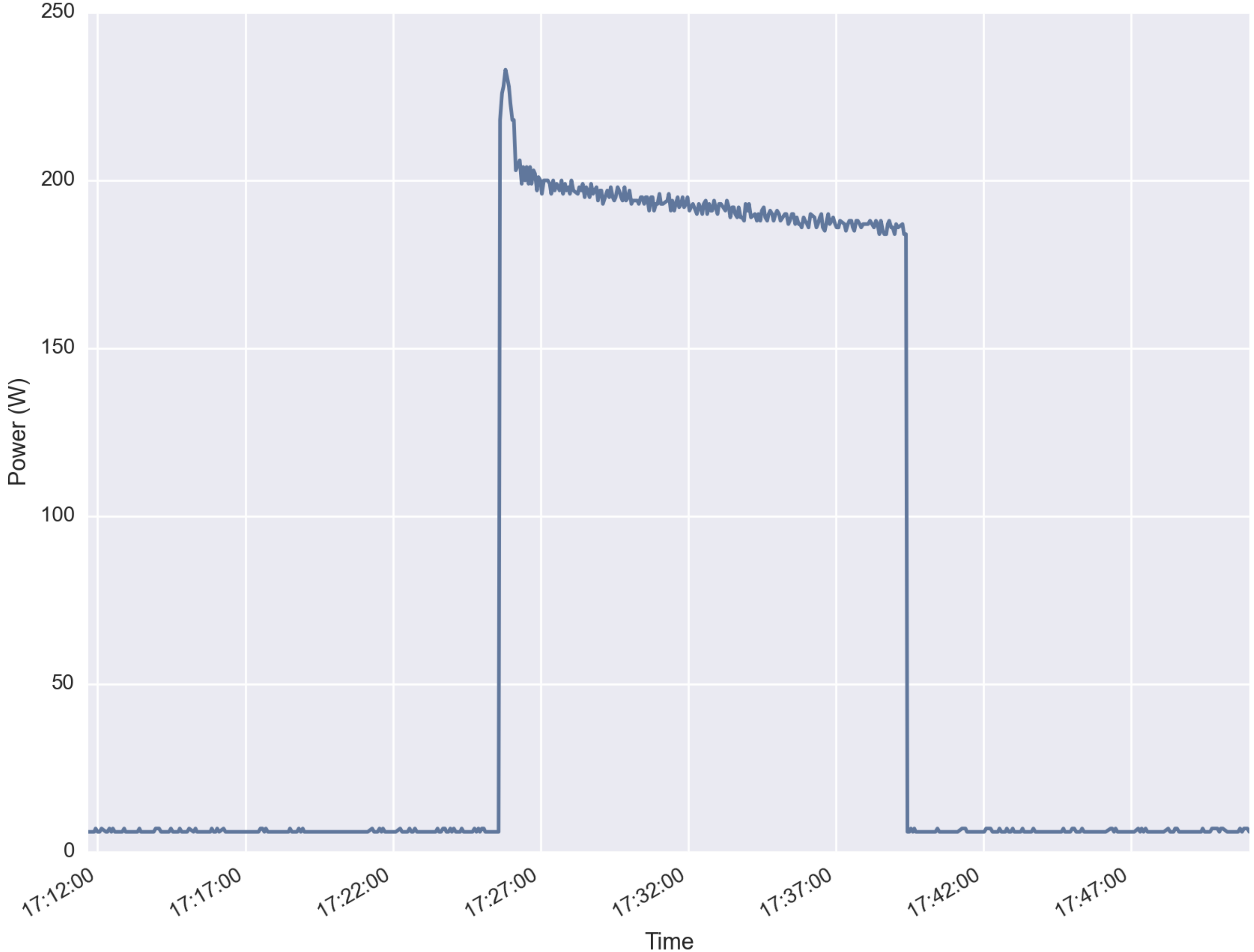

60
50
40
30
20
10
0
Houses
0.0
0.2
0.4
0.6
0.8
1.0
Proportion of energy

Challenge: blind disaggregation

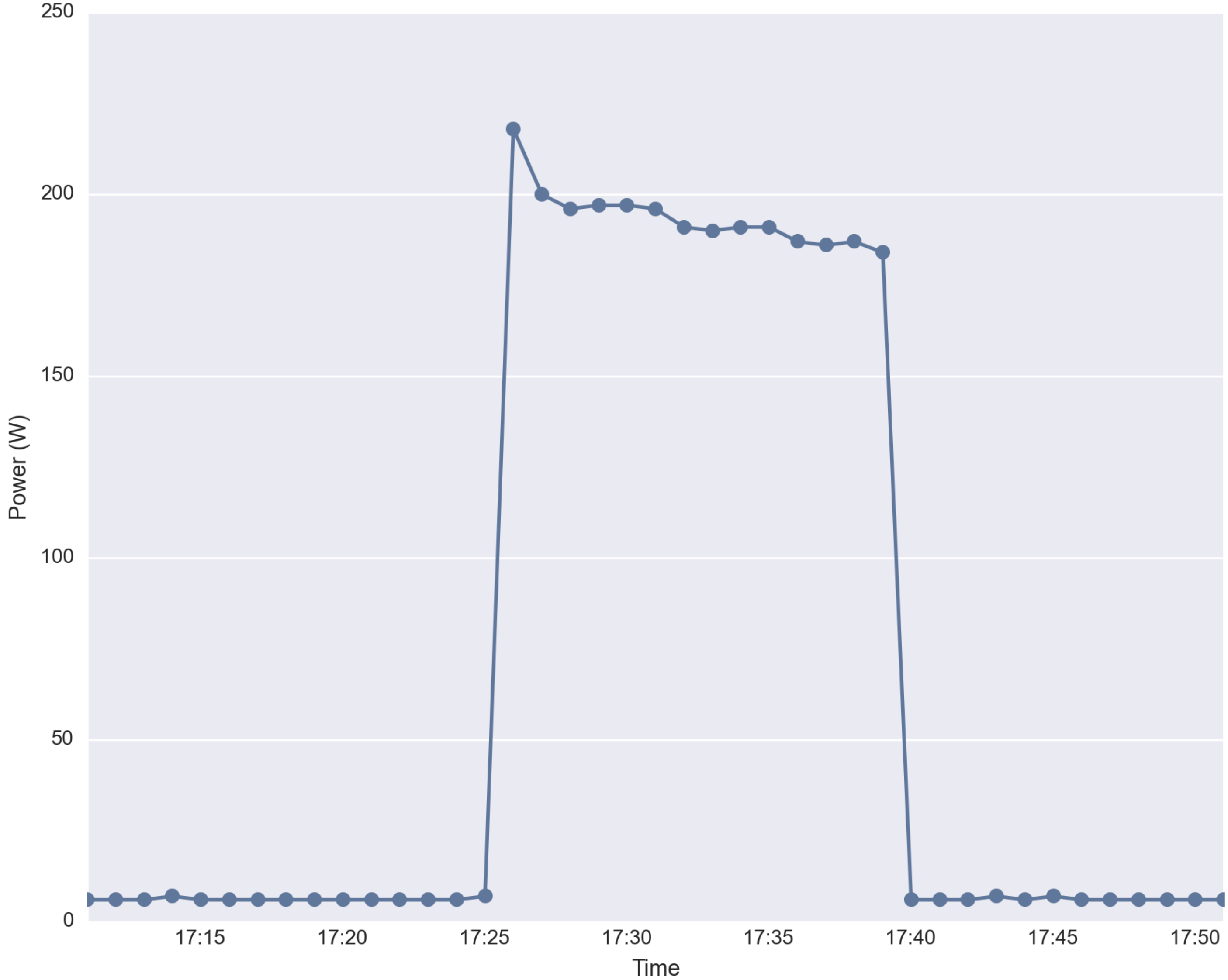
Not all appliances were created equally either…



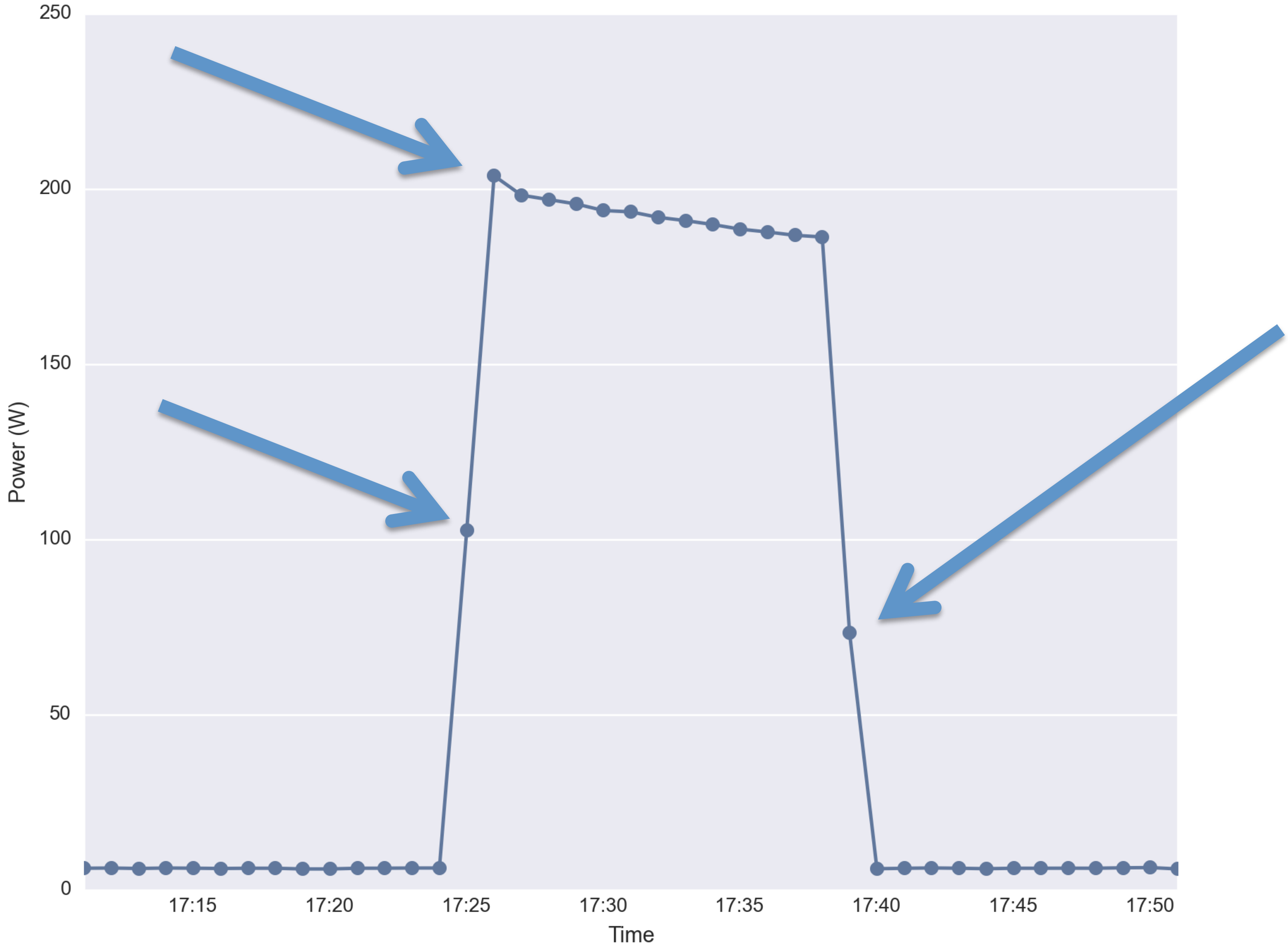
3 second power measurements



1 minute power measurements



1 minute energy measurements



Conclusion

- Dataport is the world's largest source of disaggregated energy data (and it's free*)
- The NILMTK HDF5 format makes this data set easy to use for energy disaggregation
- Dataport + NILMTK makes it very easy to investigate blind disaggregation algorithms

`https://dataport.pecanstreet.org/data/database?hdf5`